**Fake News Detection System**

**Arnab Karmakar**

**arnab.karmakar2000@gmail.com**

**9721866757,6392315115**

**Abstract**

**What is Fake News?**

A type of yellow journalism, fake news encapsulates pieces of news that may be hoaxes and is generally spread through social media and other online media. This is often done to further or impose certain ideas and is often achieved with political agendas. Such news items may contain false and/or exaggerated claims, and may end up being viralized by algorithms, and users may end up in a filter bubble.

In our modern era where the internet is ubiquitous, everyone relies on various online resources for news. Along with the increase in the use of social media platforms like Facebook, Twitter, etc. news spread rapidly among millions of users within a very short span of time. The spread of fake news has far-reaching consequences like the creation of biased opinions to swaying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to generate revenue using advertisements via click-baits. In this paper, we aim to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning. We aim to provide the user with the ability to classify the news as fake or real and also check the authenticity of the website publishing the news.

**Keywords:**

Spam Detection

Benchmark Dataset

[TfidfVectorizer](https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html" \t "_blank)

**IDF (Inverse Document Frequency)**

### Passive Aggressive Classifier

1. **Introduction**

The rise of fake news highlighted not only the dangers of the effects of fake news but also the challenges presented when attempting to separate fake news from real news. Fake news may be a relatively new term but it is not necessarily a new phenomenon. Fake news has technically been around at least since the appearance and popularity of one-sided, partisan newspapers in the 19th century. However, advances in technology and the spread of news through different types of media have increased the spread of fake news today. As such, the effects of fake news have increased exponentially in the recent past and something must be done to prevent this from continuing in the future. I have identified the three most prevalent motivations for writing fake news and chosen only one as the target for this project as a means to narrow the search in a meaningful way. The first motivation for writing fake news, which dates back to the 19th century one-sided party newspapers, is to influence public opinion. The second, which requires more recent advances in technology, is the use of fake headlines as click bait to raise money. The third motivation for writing fake news, which is equally prominent yet arguably less dangerous, is satirical writing. While all three subsets of fake news, namely, click bait, influential, and satire, share the common thread of being fictitious, their widespread effects are vastly different. As such, this paper will focus primarily on fake news as defined by politifact.com, “fabricated content that intentionally masquerades as news coverage of actual events.” This definition excludes satire, which is intended to be humorous 8 and not deceptive to readers. Most satirical articles come from sources like “The Onion“, which specifically distinguish themselves as satire. Satire can already be classified, by machine learning techniques according to. Therefore, our goal is to move beyond these achievements and use machine learning to classify, at least as well as humans, more difficult discrepancies between real and fake news. The dangerous effects of fake news, as previously defined, are made clear by events such as in which a man attacked a pizzeria due to a widespread fake news article. This story along with analysis from provides evidence that humans are not very good at detecting fake news, possibly not better than chance. As such, the question remains whether or not machines can do a better job. There are two methods by which machines could attempt to solve the fake news problem better than humans. The first is that machines are better at detecting and keeping track of statistics than humans, for example it is easier for a machine to detect that the majority of verbs used are “suggests” and “implies” versus, “states” and “proves.” Additionally, machines may be more efficient in surveying a knowledge base to find all relevant articles and answering based on those many different sources. Either of these methods could prove useful in detecting fake news, but we decided to focus on how a machine can solve the fake news problem using supervised learning that extracts features of the language and content only within the source in question, without utilizing any fact checker or knowledge base. For many fake news detection techniques, a “fake” article published by a trustworthy author through a trustworthy source would not be caught. This approach would combat those “false negative” classifications of fake news. In essence, the task would be equivalent to what a human faces when reading a hard copy of a newspaper article, without internet access or outside knowledge of the subject (versus reading something online where he can simply look up relevant sources). The machine, like the human in the coffee shop, will have only access to the words in the article and must use strategies that do not rely on blacklists of authors and sources. The current project involves utilizing machine learning and natural language processing techniques to create a model that can expose documents that are, with 9 high probability, fake news articles. Many of the current automated approaches to this problem are centered around a “blacklist” of authors and sources that are known producers of fake news. But, what about when the author is unknown or when fake news is published through a generally reliable source? In these cases it is necessary to rely simply on the content of the news article to make a decision on whether or not it is fake. By collecting examples of both real and fake news and training a model, it should be possible to classify fake news articles with a certain degree of accuracy. The goal of this project is to find the effectiveness and limitations of language-based techniques for detection of fake news through the use of machine learning algorithms including but not limited to convolution neural networks and recurrent neural networks. The outcome of this project should determine how much can be achieved in this task by analyzing patterns contained in the text and blind to outside information about the world. This type of solution is not intended to be an end-to end solution for fake news classification. Like the “blacklist” approaches mentioned, there are cases in which it fails and some for which it succeeds. Instead of being an end-to-end solution, this project is intended to be one tool that could be used to aid humans who are trying to classify fake news. Alternatively, it could be one tool used in future applications that intelligently combine multiple tools to create an end-to-end solution to automating the process of fake news classification.

1. **Background study**

Mykhailo Granik in their paper shows a simple approach for fake news detection using naive Bayes classifier. This approach was implemented as a software system and tested against a data set of Facebook news posts. They were collected from three large Facebook pages each from the right and from the left, as well as three large mainstream political news pages (Politico, CNN, ABC News). They achieved classification accuracy of approximately 74%. Classification accuracy for fake news is slightly worse. This may be caused by the skewness of the dataset: only 4.9% of it is fake news. Himank Gupta gave a framework based on different machine learning approach that deals with various problems including accuracy shortage, time lag (BotMaker) and high processing time to handle thousands of tweets in 1 sec. Firstly, they have collected 400,000 tweets from HSpam14 dataset. Then they further characterize the 150,000 spam tweets and 250,000 non- spam tweets. They also derived some lightweight features along with the Top-30 words that are providing highest information gain from Bag-of-Words model. 4. They were able to achieve an accuracy of 91.65% and surpassed the existing solution by approximately18%. Marco L. Della Vedova first proposed a novel ML fake news detection method which, by combining news content and social context features, outperforms existing methods in the literature, increasing its accuracy up to 78.8%. Second, they implemented their method within a Facebook Messenger Chabot and validate it with a real-world application, obtaining a fake news detection accuracy of 81.7%. Their goal was to classify a news item as reliable or fake; they first described the datasets they used for their test, then presented the content-based approach they implemented and the method they proposed to combine it with a social-based approach available in the literature. The resulting dataset is composed of 15,500 posts, coming from 32 pages (14 conspiracy pages, 18 scientific pages), with more than 2, 300, 00 likes by 900,000+ users. 8,923 (57.6%) posts are hoaxes and 6,577 (42.4%) are non-hoaxes. Cody Buntain develops a method for automating fake news detection on Twitter by learning to predict accuracy assessments in two credibility-focused Twitter datasets: CREDBANK, a crowd sourced dataset of accuracy assessments for events in Twitter, and PHEME, a dataset of potential rumors in Twitter and journalistic assessments of their accuracies. They apply this method to Twitter content sourced from BuzzFeed’s fake news dataset. A feature analysis identifies features that are most predictive for crowd sourced and journalistic accuracy assessments, results of which are consistent with prior work. They rely on identifying highly retweeted threads of conversation and use the features of these threads to classify stories, limiting this work’s applicability only to the set of popular tweets. Since the majority of tweets are rarely retweeted, this method therefore is only usable on a minority of Twitter conversation threads. In his paper, Shivam B. Parikh aims to present an insight of characterization of news story in the modern diaspora combined with the differential content types of news story and its impact on readers. Subsequently, we dive into existing fake news detection approaches that are heavily based on text-based analysis, and also describe popular fake news datasets. We conclude the paper by identifying 4 key open research challenges that can guide future research. It is a theoretical Approach which gives Illustrations of fake news detection by analyzing the psychological factors.

1. **Proposed Methodology**

This system which is developed in three parts. The first part is static which works on machine learning classifier. I studied and trained the model with 4 different classifiers and chose the best classifier for final execution. The second part is dynamic which takes the keyword/text from user and searches online for the truth probability of the news. The third part provides the authenticity of the URL input by user. In this paper, we have used Python and its Sci-kit libraries. Python has a huge set of libraries and extensions, which can be easily used in Machine Learning. Sci-Kit Learn library is the best source for machine learning algorithms where nearly all types of machine learning algorithms are readily available for Python, thus easy and quick evaluation of ML algorithms is possible. We have used Django for the web based deployment of the model, provides client side implementation using HTML, CSS and JavaScript. We have also used Beautiful Soup (bs4), requests for online scrapping.

What is a TfidfVectorizer?

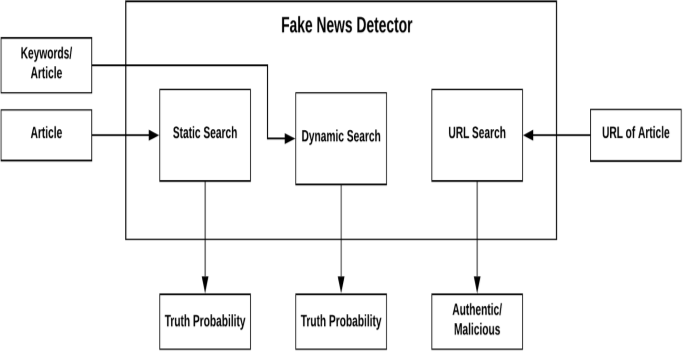
**TF (Term Frequency):** The number of times a word appears in a document is its Term Frequency. A higher value means a term appears more often than others, and so, the document is a good match when the term is part of the search terms.

**IDF (Inverse Document Frequency):** Words that occur many times a document, but also occur many times in many others, may be irrelevant. IDF is a measure of how significant a term is in the entire corpus.

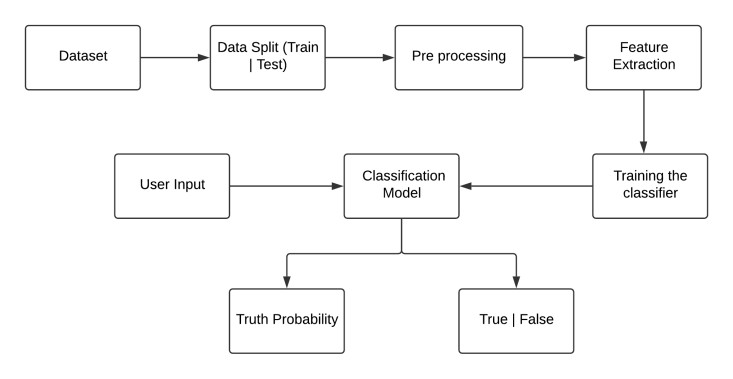
The TfidfVectorizer converts a collection of raw documents into a matrix of TF-IDF features.

What is a PassiveAggressiveClassifier?

Passive Aggressive algorithms are online learning algorithms. Such an algorithm remains passive for a correct classification outcome, and turns aggressive in the event of a miscalculation, updating and adjusting. Unlike most other algorithms, it does not converge. Its purpose is to make updates that correct the loss, causing very little change in the norm of the weight vector.

1. **System Design-**
2. **System Architecture-**
3. Static Search-

The architecture of Static part of fake news detection system is quite simple and is done keeping in mind the basic machine learning process flow. The system design is shown below and self- explanatory. The main processes in the design are-



1. Dynamic Search-

The second search field of the site asks for specific keywords to be searched on the net upon which it provides a suitable output for the percentage probability of that term actually being present in an article or a similar article with those keyword references in it.

1. URL Search-

The third search field of the site accepts a specific website domain name upon which the implementation looks for the site in our true sites database or the blacklisted sites database. The true sites database holds the domain names which regularly provide proper and authentic news and vice versa. If the site isn’t found in either of the databases then the implementation doesn’t classify the domain it simply states that the news aggregator does not exist.

1. **Implementation**

Make necessary imports:

import numpy as np

import pandas as pd

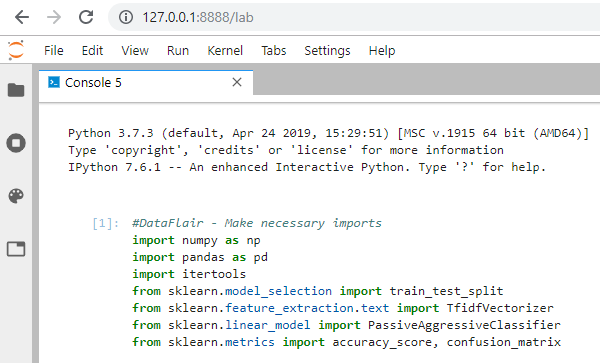
import itertools

from sklearn.model\_selection import train\_test\_split

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.linear\_model import PassiveAggressiveClassifier

from sklearn.metrics import accuracy\_score, confusion\_matrix



## Importing libraries and dataset

import re

import nltk

from nltk.stem import PorterStemmer

from nltk.corpus import stopwords

import pandas as pd

import numpy as np

import missingno

import seaborn as sns

import matplotlib.pyplot as plt

import tensorflow

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Embedding

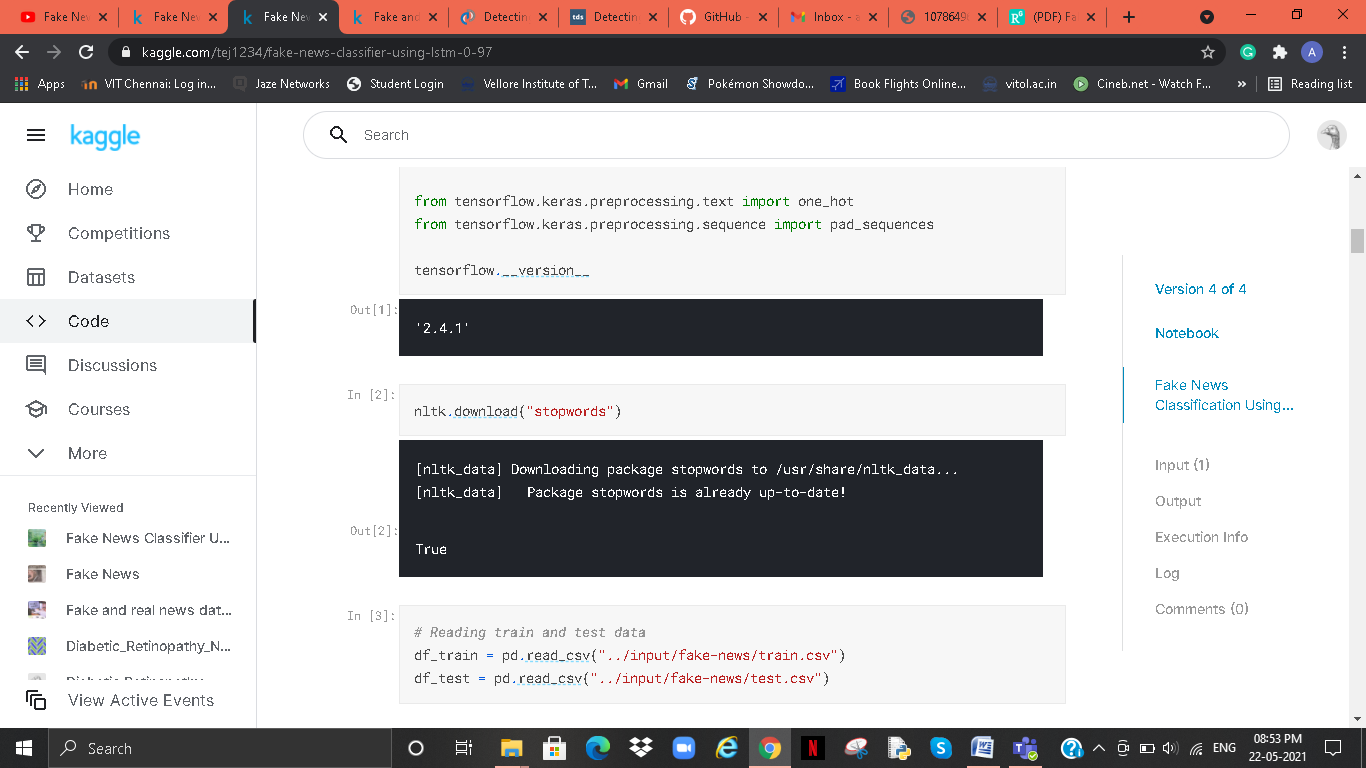
from tensorflow.keras.layers import Dense

from tensorflow.keras.layers import Flatten

from tensorflow.keras.layers import LSTM

from tensorflow.keras.preprocessing.text import one\_hot

from tensorflow.keras.preprocessing.sequence import pad\_sequences

'2.4.1'

True

1. **Results and Discussion**

Results showing testing, metrics taken for checking and a complete analysis on the results obtained.

1. **Conclusion and Future works**

The main contribution of this project is support for the idea that machine learning could be useful in a novel way for the task of classifying fake news. Our findings show that after much pre-processing of relatively small dataset, a simple CNN is able to pick up on a diverse set of potentially subtle language patterns that a human may (or may not) be able to detect. Many of these language patterns are intuitively useful in a humans manner of classifying fake news. Some such intuitive patterns that our model has found to indicate fake news include generalizations, colloquialisms and exaggerations. Likewise, our model looks for indefinite or inconclusive words, referential words, and evidence words as patterns that characterize real news. Even if a human could detect these patterns, they are not able to store as much information as a CNN model, and therefore, may not understand the complex relationships between the detection of these patterns and the decision for classification. Furthermore, the model seems to be relatively unphased by the exclusion of certain “giveaway” topic words in the training set, as it is able to pick up on trigrams that are less specific to a given topic, if need be. As such, this seems to be a really good start on a tool that would be useful to augment humans ability to detect Fake News. Other contributions of this project is include the creation of a dataset for the task and the creation of an application that aids in the visualization and understand42 ing of the neural nets classification of a given body text. This application could be a tool for humans trying to classify fake news, to get indications of which words might cue them into the correct classification. It could also be useful in researchers trying to develop improved models through the use of improved and enlarged datasets, different parameters, etc. The application also provides a way to see manually how changes in the body text affect the classification.

**Future Work**

Through the work done in this project, we have shown that machine learning certainly does have the capacity to pick up on sometimes subtle language patterns that may be difficult for humans to pick up on. The next steps involved in this project come in three different aspects. The first of aspect that could be improved in this project is augmenting and increasing the size of the dataset. We feel that more data would be beneficial in ridding the model of any bias based on specific patterns in the source. There is also question as to whether or not the size of our dataset is sufficient. The second aspect in which this project could be expanded is by comparing it to humans performing the same task. Comparing the accuracies would be beneficial in deciding whether or not the dataset is representative of how difficult the task of separating fake from real news is. If humans are more accurate than the model, it may mean that we need to choose more deceptive fake news examples. Because we acknowledge that this is only one tool in a toolbox that would really be required for an end-to-end system for classifying fake news, we expect that its accuracy will never reach perfect. However, it may be beneficial as a stand-alone application if its accuracy is already higher than human accuracy at the same task. In addition to comparing the accuracy to human accuracy, it would also be interesting to compare the phrases/trigrams that a human would point out if asked what they based their classification decision on. Then, we could quantify how similar these patterns are to those that humans find indicative of fake and real news. Finally, as we have mentioned throughout, this application is only one that would be necessary in a larger toolbox that could function as a highly accurate fake news classifier. Other tools that would need to be built may include a fact detector and a stance detector. In order to combine all of these “routines,” there would need to be some type of model that combines all of the tools and learns how to weight each of them in its final decision.

**References**

* [**https://www.researchgate.net/publication/343916946\_Fake\_News\_Detection\_Using\_Machine\_Learning\_Algorithms**](https://www.researchgate.net/publication/343916946_Fake_News_Detection_Using_Machine_Learning_Algorithms)
* <https://www.kaggle.com/c/fake-news/data>
* <file:///C:/Users/Sangeeta%20Karmakar/Downloads/IJCRT2006189.pdf>
* [**https://towardsdatascience.com/detecting-fake-news-with-and-without-code-dd330ed449d9**](https://towardsdatascience.com/detecting-fake-news-with-and-without-code-dd330ed449d9)
* D. Britz. (2016, Feb) Implementing a cnn for text classification in tensorflow [Online].

http://www.wildml.com/2015/12/implementing-a-cnn-for-text-classification-in-tensorflow/

* E. Kiely and L. Robertson. (2016, Dec) How to spot fake news [Online]. <https://www.factcheck.org/2016/11/how-to-spot-fake-news/>
* A. Karpathy, “convolutional neural networks for visual recognition.” [Online]. http://cs231n.github.io/understanding-cnn/